

# علم اللغة والتّقنيات المعاصرة

د.المُعْتَزُّ بالله السَّعيد

moataz@cu.edu.eg

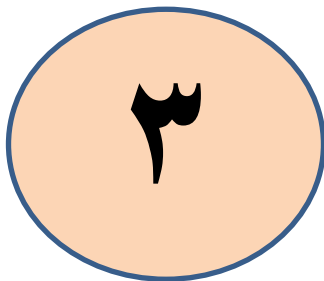
مرحبا بكم

welcome



المُحاضِرة الثَّالِثَة

فِي حَوَسِبَةِ اللُّغَةِ



- تقنيات اللُّغات الإنسانيَّة -

- من المؤسَّسات المعنِيَّة بحوسبة اللُّغة -

- أسئلة تقويمة -

- تقنيات اللُّغات الإنسانيَّة -

- من المؤسَّسات المعنيَّة بحوسبة اللُّغة -

- أسئلة تقويمة -

# تقنيات اللُّغات الإنسانيَّة



مُتَابَعَةٌ...

# ٣. تقنيات معالجة النصّ

الترجمة الآليّة

التنقيب في النصوص

المُدقّق الإملائيّ

استرجاع المعلومات

الإجابة الآليّة عن الأسئلة

فكّ الالتباس الدلاليّ

## ٣. تقنيات معالجة النصّ

### • الترجمة الآليّة Machine Translation:

✓ إجراء حاسوبيّ، يهدف إلى ترجمة النصوص بين لغتين أو أكثر.

✓ من مناهجها: منهج الترجمة المباشرة، ومنهج الترجمة التحويلي [يتم فيه ترجمة مفردات الجملة، ثمّ تحويل الناتج إلى جملةٍ مُوافقة لطبيعة النظام التركيبيّ في اللغة الهدف]، ومنهج الترجمة عبر اللغة الوسيطة [إنترلينغو Interlingua]، ومنهج الترجمة الآليّة الإحصائيّة [تقوم فكرته على تدريب نظام مُقترح للترجمة على مجموعات كبيرة من النصوص، ثمّ معالجتها باستخدام أساليب التحليل والاستنباط الإحصائيّ].

## ٣. تقنيات معالجة النصّ

### • التنقيب في النصوص Text Mining:

✓ هو عملية تحليل واستخلاص بيانات نصيّة عالية الجودة من مجموعات كبيرة من النصوص المحوسبة بعد معالجتها وتحويلها إلى نصوص مرتبطة بسياقاتها التي وُضعت فيها.

✓ من مجالاته: تلخيص للنصوص آلياً، والتعرّف على أسماء الكائنات [أسماء الأشخاص، والمنظّمات، والأماكن، ... ونحو ذلك]، والتنقيب عن الآراء.

## ٣. تقنيات معالجة النصّ

### • المُدقّق الإملائيّ Spell-checker:

- ✓ تطبيق برمجيّ، يَستَكشفُ الأخطاء الإملائيّة الواردة في النُّصوص المحرّرة والمكتوبة رقميًّا، ويُساعدُ على تصويب هذه الأخطاء عبر اقتراح مجموعةٍ من احتمالات الصّواب.
- ✓ يُوجدُ هذا التّطبيقُ باعتبارِه برمجيّة مُستقلّة بذاتها، أو مُلحقًا بأدوات التّحرير المكتبيّة.

## ٣. تقنيات معالجة النصّ

### • استرجاع المعلومات Information Retrieval:

- ✓ مجال معرفي حاسوبي. يُستخدَم في البحث عن المعلومات التي تحويها هياكل البيانات، سواءً أكانت قواعد بيانات أم شبكات معرفيّة.
- ✓ من تطبيقاته: محرّكات البحث عن الوثائق والمستندات المخزّنة في الحواسيب أو على شبكة الويب (Google, Yahoo, Bing)، وأدوات البحث المكتبيّة.

## ٣. تقنيات معالجة النصّ

### • الإجابة الآليّة عن الأسئلة Question Answering:

- ✓ تقنية حاسوبية؛ تهدف إلى إيجاد الحلّ الأمثل [الإجابة] بصورة آليّة عن الأسئلة المطروحة في إحدى اللّغات الطّبيعيّة.
- ✓ تتمّ على أنواع محدّدة من الأسئلة، وتعتمدُ على استخلاص الإجابات من الوثائق النصّيّة أو قواعد البيانات المعدّة سلفاً.
- ✓ تُستخدمُ هذه التّقنية في التّصحيح الآليّ للامتحانات.

## ٣. تقنيات معالجة النصّ

### • فكّ الالتباس الدلالي *Word Sense Disambiguation*:

- ✓ إجراء حاسوبيّ. يُقصد به إزالة الالتباس الناتج عن التداخل بين المعاني الحقيقيّة والمعاني المجازيّة لكلمة ما، استنادًا إلى السّياقات اللّغويّة الّتي تردّ فيها الكلمة موضع الالتباس.
- ✓ [عين: للإبصار، بئر، جاسوس، ...].

- تقنيات اللُّغات الإنسانيَّة -

- من المؤسَّسات المعنِيَّة بحوسبة اللُّغة -

- أسئلة تقويمة -

# مُؤَسَّسَة LDC

*Linguistic Data Consortium*

• من الموارد اللُّغَوِيَّة للعَرَبِيَّة:

✓ البنك النَّحَوِيّ العربيّ Arabic Treebank.

✓ مُعْجَم اللُّغَة الْعَرَبِيَّة الدَّارِجَة فِي مِصْر (مُعْجَم الْعَامِيَّة الْمِصْرِيَّة).

# مجموعة كولومبيا لنمذجة اللهجات العربيّة

*CADIM*

• من الموارد اللُّغويّة للعربيّة:

- ✓ نظام التّحليل اللُّغويّ MADA+TOKAN. وهو نظامٌ لتحليل النُّصوص العربيّة المكتوبة، ومن مهامّه:
- ✓ تقطيع النُّصوص Tokenization.
- ✓ التّشكيل الآليّ Diacritization.
- ✓ فكّ الالتباس الصّرفيّ Morphological Disambiguation.

# فريقُ مُعالِجة اللُّغات الطَّبِيعِيَّة، ليدز

• من الموارد اللُّغويَّة للعَرَبِيَّة:

- ✓ المدوَّنة اللُّغويَّة للعربيَّة المعاصرة، للباحثة لطيفة السُّليطي.
- ✓ المعنُون التَّلَقائِيّ لأقسام الكلام في اللُّغة العربيَّة المكتوبة، للباحث مجدي صوالحة.
- ✓ نموذج حاسوبِيّ للتَّمثِيل المعرفِيّ في القرآن الكريم، للباحث عبد الباقي شَرَف.

# المنظمة العربية للتربية والثقافة والعلوم

*ALECSO*

• من الموارد اللغوية للعربية:

✓ نظام التحليل الصرفي "الخليل".

✓ المعجم العربي التفاعلي.

✓ نظام التحليل النحوي.

✓ المدقق الإملائي للغة العربية.

✓ آلية تشكيل النصوص العربية.

# مدينة الملك عبد العزيز للعلوم والتقنية

## *KACST*

• من الموارد اللغوية للعربية:

- ✓ البنك السعودي للأصوات.
- ✓ المقوم الآلي للمقالات العربية.
- ✓ نظام معالجة الحروف العربية المطبوعة.

# الجمعية المصرية لهندسة اللغة

## *ESOLE*

• من أنشطة الجمعية:

✓ عقد مؤتمر سنوي لهندسة اللغة، يُناقش مختلف القضايا في حوسبة اللغة وتطبيقاتها.

✓ عقد ندوات وورش عمل للإفادة من خبرات المتخصصين في مجالات حوسبة اللغة.

# مركز تقنيات اللغة العربية

*ALTEC*

• من الموارد اللغوية للعربية:

✓ مجموعة من قواعد البيانات المستخدمة في التعرف على الحروف العربية – سواءً أكانت ممسوحة ضوئياً أم مكتوبةً آلياً.

✓ مدونة لغوية مضبوطة بالشكل الكلي، أنجزت باستخدام آلية التشكيل الآلي للنصوص العربية، للإفادة منها في تعيين أسماء الكائنات.

- تقنيات اللُّغات الإنسانيَّة -

- من المؤسَّسات المعنِيَّة بحوسبة اللُّغة -

- أسئلة تقويميَّة -

# أسئلة تقويمية

١. يُشارُ إلى حوسبة اللغة بالعديد من المسميات. اكتب مذكراتٍ موجزة في ذلك، ثمَّ تحدّث عن إرهاصات نشأة العلم وتطوّره.
٢. عرّف بالمصطلحات الآتية:  
(اللّسانيّات الحاسوبية ، معالجة اللّغات الطّبيعيّة ، تقنيات اللّغات الإنسانيّة)
٣. في ضوء دراستك لمُستويات التّحليل اللّغويّ في حوسبة اللّغة، اكتب مذكراتٍ موجزة حول مُستوى التّحليل:  
(الصّوتيّ ، الكتابيّ ، الصّرفيّ ، التركيبيّ ، المعجميّ ، الدّلاليّ)

# أسئلة تقويمية

٤. في ضوء دراستك لمستويات التحليل اللغوي في حوسبة اللغة، اذكر وحدات التحليل اللغوي، ثم عرّف كلّ وحدة على حدة.
٥. تقوم تقنيات اللغات الإنسانية على مستويات التحليل اللغوي عبر ثلاثة محاور. اكتب مذكرات موجزة في ذلك، ثمّ مثّل لكلّ محور من المحاور الثلاثة على حدة.
٦. من تقنيات معالجة الصورة (التعرّف على الكتابة). كيف يُستفاد من هذه التقنية في خدمة اللغة؟ وما أهم تطبيقاتها؟
٧. اكتب مذكرات موجزة حول مناهج الترجمة الآلية بين لغتين أو أكثر.

# أَسْئَلَةُ تَقْوِيمِيَّة

٨. من تقنيات اللُّغات الإنسانيَّة (التَّنْقِيب في النُّصُوص). عَرِّف بهذه التَّقْنِيَّة، ثُمَّ تحدِّث عن جوانب الإفادة منها.

٩. في ضوء دراستِكَ لتقنيات اللُّغات الإنسانيَّة، اكتب مُذَكِّراتٍ مُوجِزةً حولَ التَّقْنِيَّات الآتية:

(التَّعَرُّف على الكلام ، التَّرجمة الشَّفهيَّة الآليَّة ، البحث في الصوتيَّات والمرئيَّات ، تحويل الكلام إلى لغة الإشارة ، تحويل النِّصِّ إلى كلام ، تعلُّم النُّطق بمُساعدة الحاسوب ، التَّعَرُّف على الكتابة ، التَّرجمة الآليَّة ، التَّنْقِيب في النُّصُوص ، المُدَقِّق الإملائيّ ، استرجاع المعلومات ، الإجابة الآليَّة عن الأسئلة ، فكّ الالتباس الدِّلاليّ)



شُكْرًا لَكُمْ

# للتواصل ..

• د. المعتز بالله السعيد

[moataz@cu.edu.eg](mailto:moataz@cu.edu.eg)

[almoataz2006@skype.com](mailto:almoataz2006@skype.com)

[almo3tazbellah@facebook.com](https://www.facebook.com/almo3tazbellah)

[www.ask.fm/AlmoatazBAISaid](http://www.ask.fm/AlmoatazBAISaid)

<http://scholar.cu.edu.eg/moataz>

<http://cairo.academia.edu/AlmoatazBAISaid>